

CSE574

Introduction
to Machine
Learning

Jue Guo

Alternative
View of
Logistic
Regression

Support
Vector
Machine

Large Margin
Intuition

The
Mathematics
behind Large
Margin
Classification

Kernels

CSE574 Introduction to Machine Learning

Support Vector Machine

Jue Guo

University at Buffalo

February 3, 2024

1 Alternative View of Logistic Regression

2 Support Vector Machine

- Large Margin Intuition
- The Mathematics behind Large Margin Classification

3 Kernels

Alternative View of Logistic Regression

CSE524

Introduction
to Machine
Learning

Jue Guo

Alternative
View of
Logistic
Regression

Support
Vector
Machine

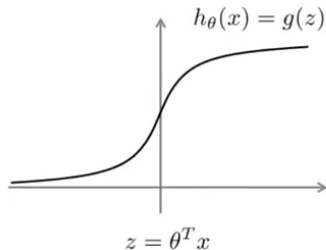
Large Margin
Intuition

The
Mathematics
behind Large
Margin
Classification

Kernels

A quick review: $h_{\theta}(x) = \frac{1}{1+e^{-\theta^T x}}$

- if $y = 1$, we want $h_{\theta}(x) \approx 1$,
 $\theta^T x \gg 0$
- if $y = 0$, we want $h_{\theta}(x) \approx 0$,
 $\theta^T x \ll 0$

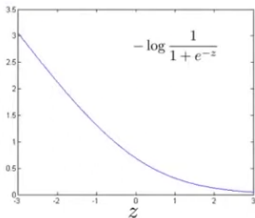


The cost of a single example:

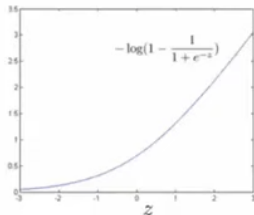
$$\begin{aligned} & - (y \log h_{\theta}(x) + (1 - y) \log (1 - h_{\theta}(x))) \\ &= - y \log \frac{1}{1 + e^{-\theta^T x}} - (1 - y) \log \left(1 - \frac{1}{1 + e^{-\theta^T x}} \right) \end{aligned}$$

$$-y \log \frac{1}{1 + e^{-\theta^T x}} - (1 - y) \log \left(1 - \frac{1}{1 + e^{-\theta^T x}} \right)$$

if $y = 1$ (want $\theta^T x \gg 0$)



if $y = 0$ (want $\theta^T x \ll 0$)



Cost Function of Logistic Regression

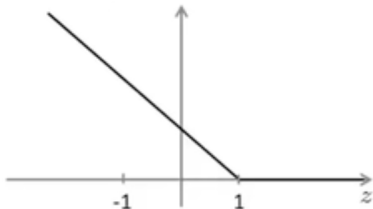
$$\begin{aligned} \min_{\theta} \frac{1}{m} & \left[\sum_{i=1}^m y^{(i)} \left(-\log h_{\theta} \left(x^{(i)} \right) \right) \right. \\ & \left. + \left(1 - y^{(i)} \right) \left(-\log \left(1 - h_{\theta} \left(x^{(i)} \right) \right) \right) \right] \\ & + \frac{\lambda}{2m} \sum_{j=1}^n \theta_j^2 \end{aligned}$$

Cost Function of Support Vector Machine

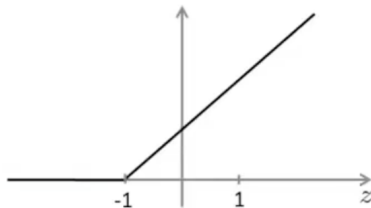
$$\min_{\theta} C \sum_{i=1}^m \left[y^{(i)} \text{cost}_1 \left(\theta^T x^{(i)} \right) + \left(1 - y^{(i)} \right) \text{cost}_0 \left(\theta^T x^{(i)} \right) \right] + \frac{1}{2} \sum_{i=1}^n \theta_j^2$$

Support Vector Machine

$$\min_{\theta} C \sum_{i=1}^m \left[y^{(i)} \text{cost}_1 \left(\theta^T x^{(i)} \right) + \left(1 - y^{(i)} \right) \text{cost}_0 \left(\theta^T x^{(i)} \right) \right] + \frac{1}{2} \sum_{j=1}^n \theta_j^2$$



If $y = 1$, we want $\theta^T x \geq 1$ (not just ≥ 0)



If $y = 0$, we want $\theta^T x \leq -1$ (not just < 0)

A very large C

Support Vector Machine

$$\min_{\theta} C \sum_{i=1}^m \left[y^{(i)} \text{cost}_1 \left(\theta^T x^{(i)} \right) + \left(1 - y^{(i)} \right) \text{cost}_0 \left(\theta^T x^{(i)} \right) \right] + \frac{1}{2} \sum_{j=1}^n \theta_j^2$$

Given that C is a very large value, we want that the first term to be 0. Let's try to understand the optimization problem in the context of what would it take to make this first term in the objective equal to 0.

Whenever $y^{(i)} = 1$, $\theta^T x^{(i)} \geq 1$;

Whenever $y^{(i)} = 0$, $\theta^T x^{(i)} \leq -1$

Now, the optimization problem can be written as:

$$\begin{aligned} \min & C \cdot 0 + \frac{1}{2} \sum_{j=1}^n \theta_j^2 \\ \text{s.t. } & \theta^T x^{(i)} \geq 1 \quad \text{if } y^{(i)} = 1 \\ & \theta^T x^{(i)} \leq -1 \quad \text{if } y^{(i)} = 0 \end{aligned}$$

SVM Decision Boundary: Linearly separable case

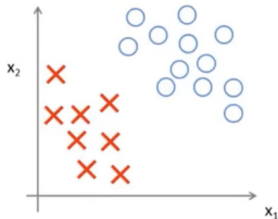
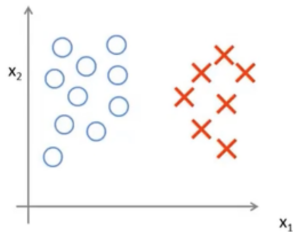


Figure: Linearly Separable Case

Large margin classifier in presence of outliers



The Mathematics behind Large Margin Classification

CSE524

Introduction
to Machine
Learning

Jue Guo

Alternative
View of
Logistic
Regression

Support
Vector
Machine

Large Margin
Intuition

The
Mathematics
behind Large
Margin
Classification

Kernels

Vector Inner Product



$$u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \quad v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

SVM Decision Boundary

CSE574

Introduction
to Machine
Learning

Jue Guo

Alternative
View of
Logistic
Regression

Support
Vector
Machine

Large Margin
Intuition

The
Mathematics
behind Large
Margin
Classification

Kernels

$$\begin{aligned} \min_{\theta} \quad & \frac{1}{2} \sum_{j=1}^n \theta_j^2 \\ \text{s.t.} \quad & \theta^T \mathbf{x}^{(i)} \geq 1 \quad \text{if } y^{(i)} = 1 \\ & \theta^T \mathbf{x}^{(i)} \leq -1 \quad \text{if } y^{(i)} = 0 \end{aligned}$$



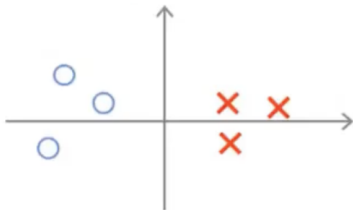
$$\min_{\theta} \quad \frac{1}{2} \sum_{j=1}^n \theta_j^2 = \frac{1}{2} \|\theta\|^2$$

$$\text{s.t.} \quad p^{(i)} \cdot \|\theta\| \geq 1 \quad \text{if } y^{(i)} = 1$$

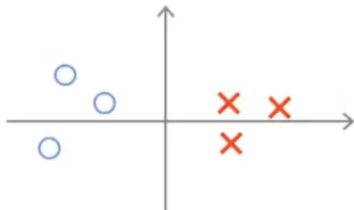
$$p^{(i)} \cdot \|\theta\| \leq -1 \quad \text{if } y^{(i)} = 0$$

where $p^{(i)}$ is the projection of $x^{(i)}$ onto the vector θ . Simplification: $\theta_0 = 0$; this simplification merely makes the decision boundary to pass through $(0,0)$;

Hypothesis 1: Bad Decision Boundary



Hypothesis 2: Good Decision Boundary



Reading Assignments

CSE574

Introduction
to Machine
Learning

Jue Guo

Alternative
View of
Logistic
Regression

Support
Vector
Machine

Large Margin
Intuition

The
Mathematics
behind Large
Margin
Classification

Kernels

1 Why the parameter vector orthogonal to the decision boundary?

- Orthogonality in Neural Network
- SVM

Non-linear Decision Boundary

CSE574

Introduction
to Machine
Learning

Jue Guo

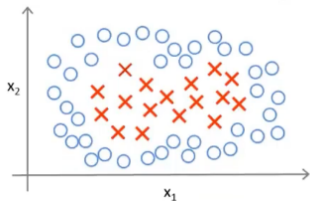
Alternative
View of
Logistic
Regression

Support
Vector
Machine

Large Margin
Intuition

The
Mathematics
behind Large
Margin
Classification

Kernels



Predict $y = 1$ if

$$\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 + \theta_4 x_1^2 + \theta_5 x_2^2 + \dots \geq 0$$

$$h_0(x) = \begin{cases} 1 & \text{if } \theta_0 + \theta_1 x_1 + \dots \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

$$\theta_0 + \theta_1 f_1 + \theta_2 f_2 + \theta_3 f_3 + \dots$$

$$f_1 = x_1, \quad f_2 = x_2, \quad f_3 = x_1 x_2,$$

$$f_4 = x_1^2, \quad f_5 = x_2^2, \dots$$

Is there a different/ better choice of the features f_1, f_2, f_3, \dots ?

Kernel

CSE574

Introduction
to Machine
Learning

Jue Guo

Alternative
View of
Logistic
Regression

Support
Vector
Machine

Large Margin
Intuition

The
Mathematics
behind Large
Margin
Classification

Kernels



Questions?